DOCUMENT RESUME

ED 189 252 UD 020 763

AUTHOR Kean, Michael H.
TITLE The Impact of Federal Evaluation Legislation on Larger Urban School Districts.
PUB DATE Sep 79
NOTE 13p.: Paper presented at the symposium on "Evaluation Policy in Federally Funded Programs: Lessons from Title I" at the annual meeting of the American Psychological Association (New York, NY, September 1-5, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Compensatory Education: Elementary Secondary Education: *Evaluation Methods: *Evaluation Needs: Federal Legislation: Federal Regulation: Policy Formation: *Program Evaluation
IDENTIFIERS Education Amendments 1974: *Elementary Secondary Education Act Title I: Pennsylvania (Philadelphia)

ABSTRACT

The Educational Amendments Act of 1974 added an evaluation requirement to Title I legislation. Models of evaluation were required to specify techniques and criteria which facilitate comparisons of programs on a statewide or nationwide basis. The School District of Philadelphia conducted simulated studies of three evaluation models to determine implementation problems and barriers regarding Title I evaluation and considered how these experiences should guide Federal policy and legislation in other programs. Four categories of barriers to smooth implementation were identified: (1) validity of evaluative measurement devices: (2) accuracy of systems for sampling, testing, and evaluation: (3) aggregation issues regarding levels at which results can or should be used: and (4) issues of interpretation and application of data. It was concluded that Title I evaluation models may be theoretically sound: however, application may lead to inconsistencies. (Author/MK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

JUL 7 - 1980

ED189252

THE IMPACT OF FEDERAL EVALUATION LEGISLATION ON LARGE URBAN SCHOOL DISTRICTS

By

Michael H. Kean, Ph.D.
Executive Director
Office of Research and Evaluation
School District of Philadelphia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Michael H. Kean

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Presented as Part of a Symposium on
"Evaluation Policy in Federally Funded Programs:
Lessons from Title I"

At the Annual Meeting of
The American Psychological Association

New York City
September 1-5, 1979

UD020763

THE IMPACT OF FEDERAL EVALUATION LEGISLATION ON LARGE URBAN SCHOOL DISTRICTS

By

Michael H. Kean, Ph.D.

INTRODUCTION

Since 1968, the evaluation of Title I projects in Philadelphia has been the responsibility of the Office of Research and Evaluation's (ORE) Division of Federal Evaluation Resource Services (FERS). The FERS staff consists of a director and approximately 20 evaluators, (research associates and assistants), who are responsible for the evaluation of individual, federally funded projects. In addition, there are secretaries and a technical support staff which provides test construction, computer programming, and data analysis services. FERS responsibilities include proposal development and needs assessment, development of objectives, data collection and analysis, and reporting.

Over the years, efforts to synthesize data from the individual state evaluation reports have been frustrated by a lack of common content and reporting practices. With the passage of the Educational Amendment Act of 1974, Congress added a federal level evaluation requirement to the legislation. The act required the Commissioner of Education to undertake certain activities with regard to Title I evaluation:

- to provide for independent evaluations
- to publish standards for project evaluation
- to consult with States regarding evaluation studies
- to provide models for evaluations of Title I projects

- to provide technical assistance in support of the systematic evaluation of programs in accordance with the models

- to disseminate information about effective educational practices.

The models were to specify techniques and criteria which would enable comparability of evaluations and, therefore, facilitate comparisons on a statewide or nationwide basis. The developmental effort to formulate a system for providing comparable data was performed under contract by RMC Research Corporation. Their efforts were aimed at establishing techniques and standards that would ultimately result in the capability for determining what benefits (at the local, state, and national levels) have accrued from the expenditures made under Title I, ESEA. Also, there will be information developed that may help in identifying successful instructional treatments.

The models were all constructed to yield a measure of treatment effect defined as the difference between observed post-treatment (posttest scores) and an estimate of what performance would have been without the special instructional treatment. The models differ primarily with respect to the manner in which this no-treatment estimate is generated. Regardless of the model used, however, cognitive achievement gains are expressed in terms of a common metric, the Normal Curve Equivalent (NCE). This unit is a normalized standard score that has been linearly transformed to coincide with the percentile distribution at values of 1, 50, and 99. The scale has a mean of 50 and a standard deviation of 21.06. These characteristics produce an equal interval scale and permit data to be aggregated from the classroom to the whole nation.

OBJECTIVES

This presentation will examine implementation problems and barriers regarding Title I evaluation in the School District of Philadelphia and consider how these experiences should guide federal evaluation policy and legislation in other programs. Four primary categories of barriers or threats to smooth implementation will be considered. The first two involve data quality, the latter two use an interpretation of evaluation in program planning and improvement of evaluation practice. Each has implications for inservice training, technical assistance, and cost required to inform teachers, principals, program administrators, and research and evaluation staff of the requirements and procedures involved in measurement and evaluation models recommended by USOE. The four categories of barriers considered are:

1. Validity Title I evaluation has focused on measurement models requiring use of standardized achievement tests matched in difficulty to ability levels of Title I participants, creating a demand that test selection and administrative procedures be reassessed.
2. Accuracy While teachers can be trained to convert test scores to the common metric required by USOE, methods of assuring accuracy and editing data suggest the need for large-scale data base systems for sampling, testing, and evaluation practice.
3. Aggregation Once data quality issues are resolved, aggregation issues arise regarding the levels at which results can or should be used in local and state reporting (e.g., classroom, grade level, building, project or activity area, district).

4. Interpretation and Application Since Title I represents extra services supplementing education funded and controlled by state and local resources, many issues remain unresolved regarding use of such results when they focus too narrowly on federal contributions alone.

BARRIERS TO IMPLEMENTATION

Validity

In focusing on validity, it is not the purpose of this section to debate the merits of the various norm-referenced standardized tests most commonly marketed. Questions of norming procedures, testing time, in- and out-of-level testing, and logistics are of interest. Of considerably more concern, however, is that the new Title I regulations represent a return to the era which equated evaluation to standardized testing. The form and substance of the proposed regulations have entirely overlooked the advances within the field, that include process and implementation evaluation. Nowhere within the proposed models are there assurances that the Title I project has been implemented as designed. Under such conditions, test scores may be measuring non-events. A related problem is that of the identification of promising compensatory programs. There are no provisions for documenting the basic operating programs which underlie Title I. How can one identify promising practices as an isolated phenomenon? Is it not possible that similar Title I services might be either effective or ineffective depending on how they interact with the regular program?

The requirement of pre- and posttesting within the same year is costly, both in terms of time and money. The informational benefits are minimal, whereas annual testing over several years can effectively provide the information concerning achievement of project participants at a more reasonable cost.

Reliability and validity of measuring devices will present problems to those who will select models A_2, B_2, C_2, using criterion referenced tests (CRT's). According to Popham,

> A criterion-referenced test should not be faulted if, when administered after instruction, everyone obtained a perfect score. Yet, that would lead to a zero internal consistence estimate, something measurement books do not recommend....it is even possible for a CRT to have a negative internal consistency index and still be a good test.

It may, therefore, be difficult, given the current technology, to prove that one is in compliance with section 116.a30 (b) if one uses models which allow a CRT testing device. Norming a CRT need not enhance its reliability.

Accuracy

Assumptions are made about the data to be provided by school staffs that simply do not reflect the realities faced by teachers and principals. It is assumed, for example, that teachers can respond to the information required on the building-level impact form, that is, the Pupil Cognitive Achievement data and the Instructor to Student Ratio.

In reality, teachers can fill in the achievement information, determine means, and compute NCE's. However, teacher involvement would necessitate special training sessions. Training of at least one teacher in each of Philadelphia's 150 Title I schools in order

that they can train their colleagues, would require release time of at least two hours or a cost in excess of $14 per hour per teacher after school. Training of all teachers to complete the forms would be at a minimum cost of over $120,000 or a loss of 8000 instructional hours. These figures do not even include additional release time or cost for actual completion of forms.

Teachers will encounter greater difficulty in responding to other impact information. For example, the impact form also requires computation of the Instructor to Student Ratio. In Philadelphia, some of our individualized Title I reading projects pay for materials. The teachers in these classrooms are paid through the operating budget. By definition, the Instructor to Student Ratio includes only teachers paid by Title I. The ratio in these instances would be 0 to 1.

Fortunately for a system as large as Philadelphia, the individual pupil achievement score can be extracted from the city-wide (central) test file. However, extraction involves accurate tracking of Title I children and the services they have received. We are in the process of finalizing the development of such a tracking system. The cost of developing such a system, however, has been substantial. It is unlikely that smaller school districts could provide the resources necessary to develop a similar approach.

It has also been assumed that building level administrators (principals) are aware of the programmatic details required on the building level forms. That is, building principals know such information as who is funded under various federal projects, what type and amount of training they receive, and the costs of various projects within

their respective building.

In reality, principals in large cities may be uncertain of the funding source of a given project. They are even less sure of the particulars of who? where? when? and how much?

In Philadelphia, computer files have available information such as PBRS Payroll Distribution, Categorical Funds Position Control Files, Project Location matrix, and Project Activity audit trails. These files may be interrogated to extract the kind of information required under the General Information Category (i.e., non-cognitive achievement data). Again, however, it is unlikely that smaller school districts have available information systems of this type.

Aggregation

There are two basic concerns relative to the use of the Title I proposed models.

1. There is no empirical evidence that the models are comparable and will provide similar results. A recent local study of this problem indicated that the findings and conclusions were a function of the selected model. This was based on a post hoc study which used the same children in the treatment groups across all three models. (Davidoff and Iwamoto, 1978).
2. As far as can be ascertained, there are no provisions for determining how Title I works in concert with the "regular program." How can one begin to make an estimate of performance in the absence of Title I?

The above points go to the heart of the matter in terms of the models and the reporting system. Further testing is a fundamental necessity. As a case in point, the study cited above might be examined in more detail. This study was predicated upon the assumption that the results obtained from the use of Models A, B, C are comparable.

In order to test the comparability of results obtained through the use of Models A, B, and C, twelve Title I elementary schools which participated in the Elementary Mathematics Resource Teacher project during the 1976-1977 school year were selected. The scores obtained from the Concepts and Problems section of the California Achievement Tests (CAT-70) were used as the selection criteria for Models A and B, and as the pretest measure for Model C. The February 1976 and February 1977 Computation scores were used as the pretest and posttest measures respectively for Models A and B. The February 1977 Computation scores were used as the posttest measure for Model C. The same 417 students in grades 2-6 served as the treatment group in the simulated runs of Models A, B, and C.

Model A: Overall, the students in the treatment group achieved a weighted mean NCE gain of 2.18. Positive NCE gains were found for grades 3, 5, and 6; negative gains were found for grades 2 and 4.

Model B: The posttest scores of students in the treatment group were compared to the scores of 382 students who were eligible, but who did not receive Title I remediation service. Grade 2 scores did not require an adjustment of the posttest mean (i.e., difference between pretest NCEs was 0). Grade 4 scores were excluded from the analysis because they did not meet the assumptions underlying

the covariance adjustment procedures (i.e., linearity). The weighted mean NCE gain obtained from grades 2, 3, 5 and 6 was 3.34. All grades showed positive gains.

Model C: The scores of 4,029 students who had scored above the 16th percentile on the Concepts and Problems section of the CAT-70 in February 1976 were compared to the treatment group. The scores of only the grade 5 students met the assumptions underlying regression analysis (i.e., linearity). They showed a mean NCE gain of zero.

Although this study was performed on a post-hoc basis, the treatment group was composed of the same students across all three models. Did Title I have an impact? According to

Model A: Yes -- Title I service had an overall positive impact. All grades were included in the analysis. Positive gains were reported for grades 3, 5, and 6.

Model B: Yes -- Title I service had an overall positive impact. Grade 4 scores were not included in the analysis. Positive gains were reported for all grades that could be included in the analysis.

Model C: No -- Title I service had no overall impact. Grades 2, 3, 4 and 6 scores were not included in the analysis. Grade 5 showed no gain. For implementation of this model, a larger number of students than originally suggested (i.e., the manual suggests N=30) may be necessary before the assumptions can be met.

Interpretation and Application

It has been assumed (erroneously) that impact information which relates student achievement data with project exposure data, costs, etc., will permit the identification of promising compensatory practices.

In reality, Title I was designed to provide extra services, meant to be used in addition to state and local resources. An extra layer, as it were.

As far as can be determined, there are no provisions for documenting the basic operating programs which underlie Title I projects. How can one identify promising practices as an isolated phenomenon, especially if they are designed to work in concert with a basic program?

Is it not possible that similar Title I services might be either effective or ineffective depending on how they interact with the regular program?

If we are to identify promising practices we must determine how Title I interacts with regular oeprating budget services. Therefore, provisions would have to be made for collection and categorization of the basic programs that underlie Title I..

CONCLUSION

Theoretically, the Title I Models may be sound. However, applications in the real world may lead to inconsistencies such as the results reported herein. Strict adherence to the assumptions underlying Models B and C may lead to exclusion of findings for several grades. Ignoring assumptions may be equally dangerous. Before, any aggregation of data is made, there

appears to be a serious need for further testing of all three models.

The Models, as well as other major Title I concerns, are not unique to the School District of Philadelphia. It is our hope that the proposed regulations can be further modified to take them into consideration. In this way, we may become partners in conducting evaluations that will lead to program improvement and help our children progress from where they are to where they ought to be.